Charge:

The P20 Data Warehouse group will review the issues surrounding the development of a data warehouse that is composed of data from the KEN partners and other relevant agencies for policy and research purposes. The group will evaluate all of the likely participants in the warehouse, discuss legal and policy issues surrounding data matches, assess the preparation needed by each agency to participate, make recommendations for system and business intelligence functionality, and establish a basic timeline for the project.

Scope:

SIF (School Interoperability Framework), PESC (Postsecondary Electronic Standards Council), DQC (the Data Quality Campaign), and KEDA (Kentucky Enterprise Data Architecture)

Membership

The P20 work group was composed of membership from the various agencies that were likely to be affected by or identified as partners in the development of a P20 warehouse.

CPE	Charles McGrew	Cabinet	Ann Riggs
	Dennis Taulbee		Chris Cross
KDE	Robert Hackworth	EPSB	Scott Smith
	Linda Pittenger		
	Kay Kennedy	UK	Roger Sugarman
	Lee Muncy		Steve Clements
COT	Glenn Thomas	GearUp	Yvonne Lovell
KCTCS	Erin Wright	KYAE	Marilyn Lyons
	Alecia Crouch		
KHEAA	Mel Letteer		

1. Evaluate the likely participants in a P20 warehouse project.

The primary agencies identified as members of this work group (CPE, KDE, EPSB, KHEAA, KYAE, COT, and the Education Cabinet), are vital players in an initiative of

this type. In addition, the group discussed and decided that someone should represent the independent or private college sector. When a final work group is established to work on this project, membership should also include someone from the Association of Independent Kentucky Colleges and Universities and the Proprietary Education Board.

2. Discuss legal and policy issues surrounding data matches.

The Family Education Rights and Privacy Act (FERPA) is the likely law that would need to be reviewed in depth before and during the development of a P20 warehouse. FERPA is identified as section 444 of the General Education Provisions Act, 20 U.S.C. § 1232g. Section 99.35 of the regulations states the following:

> "a) The officials listed in 99.31 (a) (3) may have access to education records in connection with an audit or evaluation of Federal or State supported educational programs, or for the enforcement of or compliance with Federal legal requirements which relate to those programs."

The section goes on to state that information collected under paragraph (a) must be protected in a manner that does not permit personal identification of individuals by anyone except the officials referred to in paragraph (a) of this section. It also states that information collected under these provisions must be destroyed when no longer needed for the purposes listed in paragraph (a) of this section.

The Data Quality Campaign commissioned a law firm to evaluate FERPA in terms of clearing up what states can and cannot do in terms of the creation of state-level P20 type systems. The general consensus was consistent with the belief that state level agencies may share data for improving instruction and evaluating policies and programs. (http://www.hklaw.com/content/%5CWhitepapers%5CFERPA_and_Longitudinal_Data_Systems.pdf

Further, the Kentucky letter from the Family Policy Compliance Officer, intended to clarify the role of the state (the CPE in this example) in terms of requiring information to be submitted and their ability to re-disclose it, indicates that information can be shared between educational agencies (just not institutions). It continues to state the importance of confidentiality and security. Clearly though, for improving education and evaluation of federal or state programs, this information can be shared between agencies. One of the provisions of FERPA as listed above, does require an expiration date for the data collected at the identifiable, unit level. This would have to be taken into consideration in the warehouse design as this data should not be kept in an identifiable form indefinitely.

Additional legal restrictions may be imposed depending upon the depth of employment information needed from EPSB's systems and the depth of information collected through the financial aid process with KHEAA. It was also identified that some data could fall under HIPAA if health related information was utilized (immunization, etc.)

Through IPEDS (Integrated Postsecondary Education Data System), the federal government was discussing implementing a national student unit record (SUR) system. Ultimately, this project was put off in part due to legislation which would effectively make it impossible for the government itself to collect and maintain all of this identifiable data. There has been some continued interest in revisiting this issue at the federal level. Independent groups such as NCHEMS have been conducting research projects (in part funded by Lumina and others) to ascertain if voluntary state-to-state projects could be developed that would circumvent the need for a national system. Should a national level system ever become a reality, the presence of a single system that could submit records from all institutions may be beneficial. However, due to privacy and other concerns such as the fear of a "big brother" data system, it currently seems unlikely that a SUR will be developed anytime in the near future.

Legal staff from all of the affected agencies must be involved at various points in the conception and development stages of building a P20 system.

3. Assess the preparation needed by each agency to participate.

KDE is currently developing the Kentucky Instructional Data System (KIDS), which will serve as their data warehouse at the state level. Information will roll up from individual districts to create a state-level warehouse that can be used as a central repository for data. It would be most effective for a P20 system to communicate with or receive data from KIDS. As KIDS does not have provisions for attaching to another system such as a P20 warehouse, once the P20 system is designed, the specifications for how it receives data may necessitate additional development on KIDS to ensure smooth communication.

CPE is currently developing KPEDS (Kentucky Postsecondary Education Data System), a longitudinal, relational data warehouse to replace their system of flat files within the Council's Comprehensive Database. It will include unit-level information from all of the public and independent (4-year, not-for-profit member institutions of AIKCU) colleges and universities. As the central repository of information from the institutions, it would be most efficient if this single system communicated with the P20 warehouse to provide data from postsecondary education.

EPSB, KHEAA, and others will also require development to participate. Funding may be needed for this purpose at the agency level or when the system is developed, it should incorporate provisions to fund these initiatives.

4. Make recommendations for system and business intelligence functionality.

During discussions, it became apparent that there are two basic ways of achieving P20 data warehouse functionality:

1. Develop a warehouse that is centrally located and houses a subset of records extracted from of the agencies. People would go to this location or system to get reports and develop their own queries for analysis purposes.
2. Develop a warehouse model that acts as a central point for sending requests as needed to the affected partner systems and then collecting these data together to provide the needed information. No actual data would be stored in a central location. It would effectively go to each system needed to collect the pieces it needed from each and compile them "on the fly" based on the question or need.

Both have advantages and disadvantages that go beyond the scope of what this group can recommend. When the project begins, however, both should be evaluated to determine which is more appropriate given the mission of the project.

Due to legal restrictions (FERPA), the system must be developed with security in mind. While agencies may share data under FERPA, they may not disclose personally identifiable data outside of the agency level. In addition, as educational records contain a significant amount of confidential information, it is important to construct it in a way that protects individual anonymity. One of the first issues that must be overcome (this has also been identified by the Data Quality Campaign as a crucial first step), an education identification number must be developed that follows students across agencies. Without a proper key such as this, data cannot be merged at the individual level.

Standards exist currently in the form of SIF and PESC at the K12 and postsecondary level. These individual standards groups are currently working together to ensure compatibility where they overlap. In addition, COT has groups that are establishing standards of their own in terms of business intelligence and data architecture. Aside from encouraging the adoption of SIF and PESC standards, the work group encourages the P20 project to follow the standards identified by the Interoperability work group.

The final purpose of the P20 data warehouse must be determined in terms of the audience before business intelligence, reporting, and analysis recommendations can be finalized. The agencies and their staff have a different analytical capacity to interact with such a system than the general public, if the system were to function as the backend to a web based or other publicly accessible system where individuals could request reports. In general though, the system should be accessible through the internet and overall tools should be easy to use but allow for more robust queries and views as users need them. Once the project is established, it should be discussed if all of the agencies would like to funnel requests from the press, public, and LRC to this central location where possible. If so, the level of tools and importance of the interface will be considerable. Cross-agency projects such as the Council and KHEAA's affordability projects, and the Kentucky High School Feedback Report should be able to be compiled form such a system, saving individual agencies significant amounts of work.

5. Establish a basic timeline for the project.

Due to the federal reporting requirements of No Child Left Behind (NCLB), the increasing need to follow student data across agency lines for policy purposes, and the growing number of grants and other funding opportunities that require alignment of data systems, plans for a P20 system should be undertaken right away. For example, Kentucky is currently a finalist for a thirteen million dollar National Math and Science Initiative grant. If received, data about student course activity must cross from secondary to postsecondary education in order to meet the reporting requirements for evaluating the grant. The state has already said that this will occur, though no system is currently in place to do it. Work should move forward as soon as possible.

While CPE and KDE develop their new data warehouses, it would be considerably easier for them to have systems that will interface with the P20 system if those specifications were known in 2008 before both projects are completed.

The first step in the time-line would be to request planning funds to develop the overall plan for the architecture and system. This should occur in fiscal year 2008-09 and will take a complete year. Beginning in 2009, the system should be constructed. Depending on the scope of the system, it is likely to be constructed in phases. A working system should be available by the end of 2009-10, with additional functionality added during the 2010-12 biennium on an incremental basis.

Federal and external funds may be available to assist with this project development. It had been hoped that the Data Quality Campaign could be a source of funding but that does not appear to be the case. Their efforts though could help identify best practices and provide guidance, such as they have done by having an interpretation of FERPA. It is likely that such a system will be expensive, one suggestion is that a functional work group should be started right away and that they should begin looking immediately for funding to help develop such a system.

I. Where are we in the use of this application:

Currently, no P20 warehouse exists. The best examples of efforts toward this type of system would be in the limited instances where agencies are already sharing data such as the Kentucky High School Feedback Report produced as a collaborative effort between the CPE, KDE, and KHEAA.

For other specific purposes, data is shared between KDE and KHEAA and between the CPE and KHEAA. Current efforts to share student unit level data are complicated in part due to legal constraints and in part due to differences between the platforms and systems agencies are using.

II. Where we want to go in the use of this application:

As stated previously, there is a growing belief that we need to follow educational records from one level to the next in order to have the data to truly research the impact that changes in education are having in the long term. Without an understanding of how well a student is prepared for college or how well they perform in college, efforts to improve readiness in high school cannot really be evaluated.

Ultimately, the goal would be that information about student performance in K12 could be linked to that in postsecondary and that postsecondary performance and related data about prospective teachers could be evaluated by comparing it to the performance of their students once they have graduated and are in the classroom. In addition, the hope would be that professional development for teachers could be evaluated by linking it to the performance of their students.

The overall plan is to be able to provide feedback for policy research about how changes within one area affect the others.

III. How we are going to get there:

As stated in item 3, each of the respective agencies which have some of the data needed to create the P20 system will have to make accommodations so that their systems are equipped to work with such a system efficiently. This will require completion of projects at KDE and CPE currently and possibly at EPSB and KHEAA to prepare for bringing the information together.

A P20 work group will be established with representatives from the affected agencies, COT and other areas to develop the project plan and pursue funding opportunities. Funding must be established through the legislature and/or other fund sources such as federal grants. Once funding is established, project management and oversight would need to be established along with a steering group to fully evaluate the needs and expectations of such a system.

IV. Impact on teaching and learning:

A P20 system would provide unprecedented information about the outcomes of changes in the education system and truly allow for changes in teacher preparation and professional development to be measured to determine what works most effectively. It would identify where there are gaps in the curriculums between secondary, postsecondary, and even graduate study. This system would be the single best source of measuring the effects of changes in education and identifying what seems to work best.